Simulation Studies and the Design of the First GARP Global Experiment

ROBERT JASTROW and MILTON HALEM

Institute for Space Studies, Goddard Space Flight Center, NASA, New York, N.Y. 10025

Simulation studies have played an increasingly important role in planning for the First GARP Global Experiment (FGGE) during the last two years by providing information on the meteorological impact of various untested or incompletely tested elements of the GARP observing system. The studies are not concerned directly with the research objectives of GARP, but are intended to provide information to FGGE planners on the capabilities of alternative systems for meeting a specific FGGE requirement. The simulation studies also provide a valuable input to cost-effectiveness studies by indicating tradeoffs of one possible type of observing platform against another-such as constant-level balloons vs. buoys. These tradeoffs can lead to major gains in research yield per dollar of expenditure on the observing system.

1. The FGGE observing system

The observing system planned for FGGE is a mixture of conventional observing systems that are already a part of the World Weather Watch, and relatively new satellite- and balloon-based systems for the observation of winds, temperature, humidity, cloud cover, and heat budget. Some elements of the FGGE observing system, and in particular the WWW network, have been tested in years of operational use. These are standard systems whose capabilities and costs can be considered as com-

pletely firm elements in FGGE planning. Other elements have been tested on a trial basis only, and in some cases exist only as proposals.

Global satellite temperature data are an example of an important element of the FGGE observing system that has been put into limited operational use, but have not yet been tested at the level of capability required for the FGGE data requirements. Another example is the measurement of winds by tracking of clouds in satellite photographs, which also has been used on a limited basis operationally, but is still under development with major improvements in prospect. Finally, a potentially critical element, still in an early development stage, is the Carrier Balloon System, designed to satisfy the FGGE requirement for accurate vertical wind profiles in the tropics.

2. Description of the method

Simulation studies aimed at the investigation of the capability of an observing system proceed according to the following sequence of steps: First, an artificial history of the atmosphere is created by numerical integration of a model. Second, simulated "data" are created from the history by addition of random variations to the history values for temperature, wind, and pressure; these random variations represent errors in the data to be yielded by the observing system under study.

Third, the numerical integration that created the history is repeated, but with the meteorological vari-

Adapted from a report to the Joint Organizing Committee-7, Munich, June 28-July 4, 1972.

ables in the model replaced by the simulated data at locations and times correspondingly to the assumed pattern of observations.

For example, if the observing sub-system under study is designed to produce wind data, the winds in the history are replaced by simulated wind "data" at locations, heights, and times corresponding to the coverage expected from the observing system. If the study is directed at the performance of an observing sub-system designed to yield temperature or pressure information, the temperature or pressure values in the history are replaced by the respective simulated "data" in a similar fashion.

If the inserted data had no errors, and therefore were identical with the history values, the new integration would be identical with the history. However, when errors are present, the inserted data perturb the computed circulation, causing it to depart from the history. The difference between the history and the perturbed circulation resulting from the data insertion is a measure of the effect of the errors in the simulated data.²

The effect of the errors usually is expressed in terms of rms differences of the meteorological variables such as wind components, averaged over all points of the computing grid. These rms differences are considered to represent the errors in the determination of the global atmospheric states, resulting from the assumed errors in the observing system.

3. Limitations on simulation studies

An examination of the underlying rationale for FGGE simulation studies indicates important limitations on the usefulness of these studies. The most important weakness of the studies stems from the fact that the same numerical model generally is used both to generate the simulated "observations" and to test the effectiveness of these observations in a data-assimilation scheme. As a consequence, the insertion of the data produces a minimum amount of shock, and unrealistically small errors.

Even if different models are used to generate the data and subsequently to assimilate these data, the simulation study still is likely to underestimate the amount of shock produced by the data insertion, because all the models currently used in simulation studies parameterize the physics of the atmosphere in approximately the same way.

In real life, the observations are derived from the actual atmosphere, whose physics is certain to be considerably different from the physics of the models currently employed in the data assimilation program. Even if the observing system were perfect, and the observations had zero errors, the insertion of these perfect ob-

servations into the model would generate a significant level of disturbances, because of the incompatibility of the model physics with the properties of the real atmosphere.

These "incompatibility" disturbances are a measure of the defects of the model, rather than the imperfections of the observing system. However, they are just as damaging as if they were the result of observational errors. They add to the effect of the observational errors, degrading the quality of the atmospheric states yielded by the data assimilation schemes and lowering the quality of the resultant forecasts.

Thus, simulation studies on the hypothetical performance of an observing system always lead to a more favorable result than can be expected from the real observing system. Preliminary experiments with real data insertion, described below, suggest that the "incompatibility" effect contributes several meters per second to the global mean wind error. It is possible that when the data-assimilation schemes are refined, this effect will emerge as the major source of error in the global atmospheric states yielded by FGGE.

Another weakness in simulation studies relates to the problem of model-dependence. As noted above, in these studies the effectiveness of a set of observations is measured by the difference between two numerical solutions to the model, one representing the true history of the atmosphere, and the other constituting a perturbed state obtained by inserting the simulated data. The effect of minor defects in the physics of the model should cancel in forming this difference of two circulations. However, if the model has serious deficiences, and the computed circulation is severely distorted as a result, the information on errors obtained from the simulation study may be misleading. For example, regardless of the observing system under study, a model generating a sluggish circulation would produce unrealistically small wind errors.

4. Objectives of the FGGE simulation studies

The general objective of an observing system simulation study is to determine how the accuracy of global atmospheric states depends on the error limits, spacing and frequency of observations, types of variables to be measured, and other properties of the proposed system. In the FGGE studies, certain specific areas of investigation have received emphasis because they relate to the most innovative and least tested aspects of the FGGE observing system. The stressed areas are: 1) four-dimensional data assimilation schemes, aimed particularly at the use of asynoptic satellite data; 2) the relative effectiveness of various proposed systems for determining tropical winds; and 3) the importance of reference level measurements, and the optimum location of the reference level.

The report that follows is directed mainly to a review of work in these areas. The report was prepared at the request of the Working Group on Numerical Experi-

² In the actual experiments, the run with the simulated data is sometimes started from a different initial state, creating transient initial errors. The asymptotic error level after several days of integration seems to be independent of the initial state.

mentation of the Joint Organizing Committee of GARP, and summarizes the results obtained by groups conducting simulation studies at the request of the Joint Organizing Committee from 1969 through mid-1972.

5. Insertion of simulated satellite temperature data

Numerical experiments [1, 2, 3, 4] suggests that continued insertion of global temperature profiles in an atmospheric model uniquely determines winds and pressures through the relationships among meteorological variables contained in the equations of the model. As the temperature data are inserted the wind and pressure errors drop steadily, and level off at an asymptotic value after a period of time which is generally between 10 and 25 days depending on the mean magnitude of the errors in the initial state. According to Jastrow and Halem [2] the wind error decreases in proportion to the volume of temperature data assimilated daily, i.e., insertion of data every six hours produces twice as good results as insertion every twelve hours, hence two satellites with temperature sounders are twice as useful as one [2]. According to Williamson and Kasahara [3] and Williamson and Dickinson [5], however, the asymptotic level of errors is less strongly dependent on the volume of temperature data assimilated daily.

6. Synoptic vs asynoptic insertion

According to Goddard Institute for Space Studies (GISS) experiments, it does not matter whether the data are inserted synoptically at fixed intervals, or in an asynoptic four-dimensional scheme following the space-time track of the satellite which acquires the data [6]. The effect of the temperature data on winds and pressures seem to depend mainly on the amount of temperature information inserted each day, and not on the way in which the insertions are spread throughout the twenty-four hour interval. Fig. 1 compares results achieved by synoptic and asynoptic insertion of simulated temperature profiles. The asynoptic insertions followed the track of an imaginary polarorbiting satellite. Other conditions in the two experiments were identical. No difference can be seen in the wind errors yielded by the two assimilation procedures.

However, Smagorinsky [7] suggests that asynoptic 4-D assimilation should be superior to synoptic assimilation because it reduces the shock and therefore the mean error resulting from direct data insertion. Confirmation of this suggestion is found in work reported by Morel et al. [8], using simulated balloon wind data rather than temperature data. In simulated forecast studies using a one-layer barotropic, incompressible, hemispheric model, they obtained a lower mean error with asynoptic assimilation than with balanced synoptic assimilation. Fig. 2 shows their results for the error in geopotential heights.

On the other hand, a summary of very recent work by Morel, submitted for the purpose of compiling this report, indicates no difference between synoptic and asynoptic assimilation [9]. Further experiments are desirable to clarify the situation. A subjective guess is that synoptic and asynoptic assimilation will yield roughly the same results if raw data are inserted directly, but more sophisticated assimilation techniques may tip the balance in favor of the asynoptic method.

7. Need for global coverage

The coverage of data must be global or nearly global. Experiments performed by GISS and NCAR indicate that roughly 85% or more of the globe must be included in the mean daily coverage, if the temperature data are to be effective in controlling winds and pressures. In the GISS experiments, in which temperatures were inserted in an equatorial belt, the coverage had to be extended up to latitudes of 70° or more to obtain reductions in wind and pressure errors [2]. In the NCAR experiments various combinations of temperature, pressure and wind data were inserted either in the tropics or in the extra-tropics, but not in both. No error reductions were found outside the immediate region of insertion in these NCAR experiments on limited-area coverage [10].

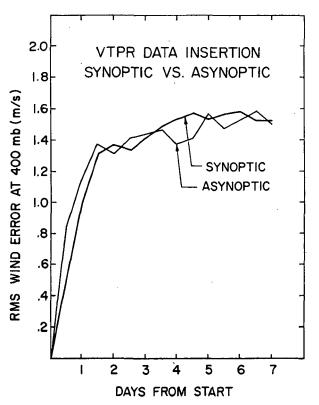


Fig. 1. Insertion of simulated VTPR sounding data (ITOS-D): comparison of synoptic versus asynoptic data assimilation results. In this case the experiments were started from unperturbed zero-error initial states; hence the wind errors rise toward their asymptotic levels, rather than decline toward the asymptotic levels from initially large values.

8. Insertion of other types of data

Similar results to those from temperature insertion are obtained if winds or pressures are inserted instead [3]. It appears that insertion of an extended time history of any single meteorological variable will determine the values of the other variable within finite error limits. It is also possible to insert redundant data—two or more fields—with similar results [6, 10].

9. The JOC tropical wind experiments

Simulation studies designed to evaluate the importance of tropical wind measurements were proposed by JOC-5 and performed between May and October of 1971 by the Geophysical Fluid Dynamics Laboratory (GFDL), GISS, and NCAR.

The studies were planned to answer the specific question: What is the minimum observing system necessary to satisfy the GARP requirement for a wind error of ±2 m sec-1 in the tropics? To this end, a sequence of three basic experiments was laid out, with major elements of the FGGE observing system successively added from one experiment to the next:

- 1) Insertion of simulated temperature profiles, representing global coverage with random and systematic errors as expected from the satellite sounding radiometers planned for FGGE.
- 2) Same as (1) plus insertion of two-level wind data in the tropical zone, simulating wind vectors derived from geostationary cloud images.
- 3) Same as (2) plus insertion of detailed vertical wind profiles within the equatorial zone, simulating dropsonde data from the Carrier Balloon System.

Curves I and II in Fig. 3 show the rms vector wind errors after 12 days as computed by Gordon, Umscheid, and Miyakoda [11] in the first experiment (temperature

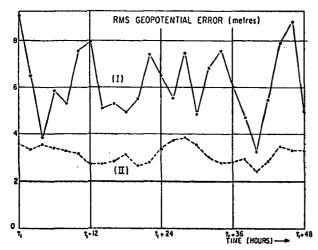


Fig. 2. Comparison of forecasts starting from (I) synoptic wind data at time t_1 (full line) or (II) asynoptic wind data and dynamic assimiliation (dashed line).

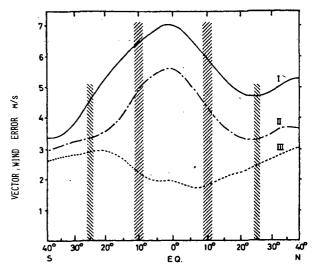


Fig. 3. Results of the GFDL tropical wind experiment, showing mean vector wind error after 12 days with the following cases of (simulated) data insertions. (1) Data inserted consist of temperature profiles with errors specified within the GARP data requirements and surface pressure. (II) In addition to I, two-level wind field (850 mb and 200 mb) with rms vector error +2 m sec⁻¹ in the lower and upper troposphere is added in the 26°N-26°S band. (III) In addition to II, complete wind profiles with rms vector error ±1 m sec⁻¹ are added in the equatorial belt 10°N-10°S.

only) and the second experiment (temperatures plus two-level winds). The important qualitative features of these results as reported by Gordon in Toronto, are:

- i) The wind errors obtained from temperature insertion alone exceed the GARP error limits.
- ii) Addition of two-level wind data in the tropics significantly improves the wind determination at all latitudes; i.e., the wind information propagates from the tropics to extratropical latitudes.
- iii) With two-level tropical wind data added, the wind errors in the tropics still exceed the GARP error limits.

With dropsonde wind data added in the equatorial zone, the wind errors in the tropics are reduced below the GARP limit of ±2 m sec-1 for tropical winds. GFDL results for this experiment are shown in curve III of the chart.

Model-dependent effects in the tropical wind experiments

Groups conducting observing system simulation studies have always been concerned with the possibility of a strong model-dependence in their results. This concern seemed to be justified when the GFDL, GISS, and NCAR groups reported their results to the JOC in Toronto in October 1971, because substantial differences were apparent between the GISS and NCAR results on the one hand and the GFDL results on the other. GISS and NCAR differed from GFDL with respect to two of the three main conclusions regarding tropical winds

that were listed in Section 9. Since all groups followed similar prescriptions in constructing their simulated data sets, their contradictory results were assumed to be the result of model-dependence.

Subsequent GISS experiments indicate that the disagreement may be primarily the result of varying vertical resolution in the three models, and not the result of differences in the physics of the models. When the GISS experiments were repeated later with a model possessing nine vertical levels—the same number as in the GFDL model—results were essentially the same as reported by GFDL, although the models were quite different in their treatments of radiative transfer and moist convection.

The agreement between the GFDL and GISS results is shown in Fig. 4 for the experiments in which simulated temperature data are compared with temperature data plus two-level wind data. The horizontal grid resolution is approximately the same (\approx 400 km) in both models. Both models are seen to yield the basic results that i) extratropical wind errors are reduced by the addition of two-level wind data in the tropics and ii) the wind errors after addition of the two-level wind data still exceed the GARP error limits. The earlier GISS study performed with a two-level model had yielded contrary results on both points.

These experiments suggest that the simulation studies performed by various groups depend strongly on the grid resolutions in each model but may not depend significantly on other features of the models. However, Miyakoda [12] indicates that differences in model physics can have an important influence on results of simulation studies performed with different models. The number of model comparisons available thus far seems too limited to settle this important question of model-dependence on Observing System Simulation Studies. This area of disagreement will be resolved gradually, as a larger body of experience accumulates in the comparison of similar experiments performed by the groups working with different models.

11. Reference level experiments

The question of reference-level measurements is connected with decisions on potentially expensive elements of the GARP Observing System, in particular, southern-hemisphere buoys and constant-level balloons. Two points are of primary interest in this regard: 1) If a reference level is needed, where should it be located? Must it be at the surface, or will 200 mb—the planned height of the constant-level balloons—be equally effective? 2) How important are reference level measurements in general?

Regarding the first question—the relative value of surface pressures vs 200-mb heights—the reports on the relevant simulation studies reach divergent conclusions. Baumhefner and Julian [13] state that the location of the reference level has little effect, while Williamson and Kasahara [3] conclude that the use of correct sur-

face pressure seems to produce a better determination of the wind field than the use of correct 12-km pressure.

The experimental results on which the last evaluation was based are shown in Fig. 5. In this figure, experiment "7" shows the results of insertion of temperature data only, experiment "6" shows the results of insertion of temperatures plus 200-mb heights, and experiment "3" shows the results of insertion of temperatures plus surface pressures. The surface pressure data improve global wind errors by about 1 m sec-1, but the 200-mb heights have no effect except for an apparent slight worsening of the results.

Recent GISS experiments related to the question of the 200-mb heights differ somewhat from the NCAR experiments reported by Kasahara and Williamson, but lead to a similar conclusion. In the first GISS experiment, temperatures were inserted globally and surface pressures were inserted over land areas only, simulating the WWW network of surface stations. In the second experiment the same data as above were inserted, and in addition, 200-mb heights were inserted in the Southern Hemisphere, simulating the coverage of the constant-level balloon system planned for FGGE. The results are shown in the table below. They indicate no appreciable effect of 200-mb heights except for a slight worsening of the wind error. These experiments also support the conclusion of Kasahara and Williamson.

	Land only	Land plu 200-mb heights
Global average (m sec-1)	4.7	4.8
Zonal average (m sec-1)		
8°S	5.3	5.7
· 24°S	3.6	3.9
38°S	3.3	3.7

Regarding the second question—the need for reference pressures in addition to those provided by the augmented WWW network—the literature is again contradictory. Smagorinsky et al. [14] report evidence that

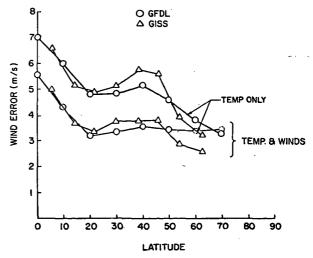


Fig. 4. Comparison between GFDL and GISS experiments on insertion of temperature plus tropical winds.

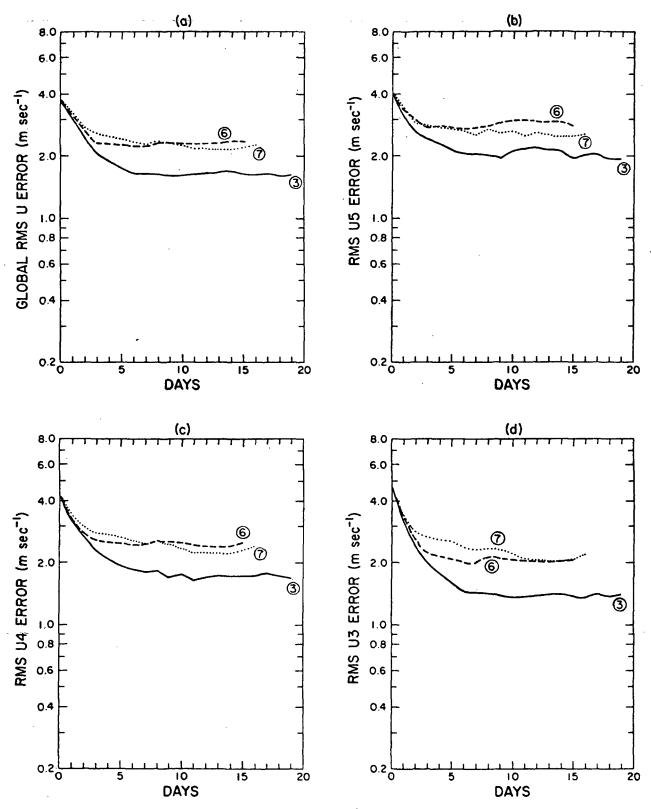


Fig. 5. NCAR reference level experiments [3]. The ordinate is the error in zonal wind (u) averaged over grid points. U3, U4, and U5 refer to levels at 7.5, 10.5, and 13.5 km, respectively.

surface pressure data "are not essential for short range forecasts in the extratropics." Kasahara states, in reporting on studies using real data, that "results of our numerical experiments under realizable observational conditions seem to suggest that the reference pressure is indeed needed" [4]. In the report by Kasahara and Williamson [10] on studies using simulated data, this statement is modified to extend to the Southern Hemisphere only:

Thus, a system of buoys over the extratropical southern oceans capable of measuring surface pressure would provide useful information for updating in the NCAR GCM. However, the observations of surface pressure over land only are sufficient for updating in the tropics and Northern Hemisphere when compared to observations of surface pressure everywhere.

Fig. 6 shows the results on which the NCAR conclusion is based. In this chart, experiment "7" shows the results of inserting temperature data only, and experiment "3" shows the results of inserting temperatures plus surface pressures globally. The dashed line shows the results of inserting temperatures globally and surface pressures over land areas only. The effect of global insertion of surface pressures, as distinct from insertion over land areas only, is seen to be small in the northern hemisphere but somewhat larger in the Southern Hemisphere. At 60°S latitude the effect is very large.

In contrast to the NCAR group, Morel finds no significant effect of surface pressure on results achieved by insertion of temperature data [9].

GISS experiments on this question were similar to the NCAR experiments and led to corresponding results. As in the NCAR results, the global mean wind error is about 1 m sec-1 larger in the "land only" case than in the case of global surface pressure insertion.

Our interpretation of the NCAR and GISS experiments that have been performed thus far on the reference-level question is the following:

- 1) A major augmentation of the WWW surface stations by a dense network of buoys with a spacing of 400 to 500 km could improve global mean wind errors by as much as 1 m sec-1.
- ii) Augmentation of the surface stations by a dense network of 200-mb balloons would have little effect.

12. The effect of inserting real data.

Morel has emphasized that simulation studies which test a model against itself will lead to unduly favorable results. In the simulation studies, the simulated "data" are obtained from a model with the same physics as the model into which these "data" are to be inserted, hence the insertion does not produce as violent a shock as results from insertion of real data. Thus the simulation studies are unrealistically optimistic.

This is the so-called "incompatibility" or "rejection" effect. In order to test it, GISS conducted a hemispheric experiment using the currently available National Meteorological Center (NMC) analyses for the Northern Hemisphere for the month of February. These provide temperatures, heights and winds down to an average latitude of about 13°. Hemispheric states were obtained by extrapolating the NMC value to the equator, using a linear extrapolation to zero for wind components, a linear extrapolation to 1010 mb for sea-level pressures, and a constant extrapolation for temperatures.

Hemispheric NMC temperatures obtained as described were inserted into the model every 12 hr, and winds derived from this data assimilation scheme were verified against the NMC winds. The rms difference in the two wind fields averaged over all vertical levels was ≈10 m sec⁻¹ (upper curve, Fig. 7).

A similar experiment was then performed using simulated temperature data generated by the model, instead of the NMC temperatures. Rms errors were introduced into the simulated data to match the random errors in the NMC temperatures. The NMC errors were estimated by comparing the NMC analyses to radiosonde measurements at the standard net of 70 stations for the same period. The coverage of the simulated data imitated the NMC coverage; the simulated data being extrapolated to the equator according to the same scheme used to extrapolate the NMC fields.

In the run with simulated data, the rms difference between the "correct" wind field and the temperature-derived wind field was found to be ≈ 8 m sec⁻¹ as compared with ≈ 10 m sec⁻¹ in the first experiment (Fig. 7). The improvement in results with the simulated data presumably is due to the fact that the model used for the data assimilation is identical to the model that was used to generate the "data."

The conclusion is that in this particular case the incompatibility effect amounted to 2 m sec⁻¹. Additional experiments will be needed to test the magnitude of the effect in a variety of conditions closer to those anticipated for FGGE.

13. Future studies

Experience over the past two years indicates two problem areas in the field of simulation studies related to FGGE. One problem area relates to the need for a better definition of certain critical elements of the FGGE Observing System, in particular, the sub-systems planned for tropical wind measurement. Another problem area relates to the inherent weaknesses in the simulation studies themselves, as an input to firm planning.

The FGGE requirement for measurement of winds in the tropics at seven vertical levels, with an accuracy of ±2 m sec-1 has emerged as the GARP data requirement most difficult to satisfy, in the sense that tropical wind measurements of this quality are farthest removed from the current capabilities of conventional or satellite observing systems. Several systems have been

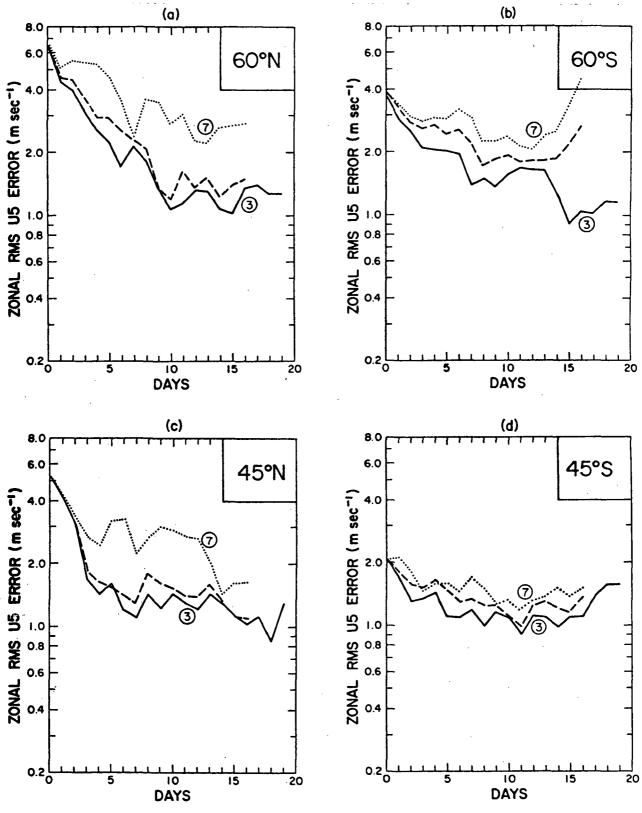


Fig. 6. NCAR reference level experiments [10]. U5 is the zonal wind at the 13.5-km level.

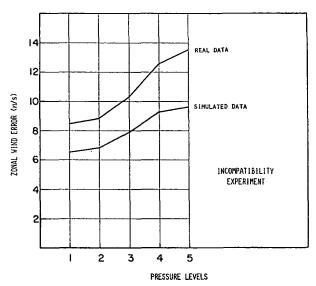


Fig. 7. Incompatibility experiment. See text for explanation.

proposed to meet the tropical wind requirement either completely or in part. These systems include the carrier balloon with multiple dropsondes, refinement of existing techniques for wind determination by cloud-tracking, and establishment of a picket line of ocean-stationed vessels for launch of rawinsondes.

The definition of the optimum system, or combination of systems, among these is one of the major tasks confronting FGGE planning in the immediate future. At an ad hoc meeting of the modeling groups concerned with GARP simulation studies, convened in New York by the Working Group on Numerical Experimentation on 12 October 1972, first priority was placed on a program of numerical experiments designed to evaluate the relative effectiveness of the above system.

Another problem that received extensive discussion at the October 12th meeting related to inherent weaknesses in the simulation studies themselves, as inputs to firm planning. It has been noted that the results obtained by NCAR, GFDL, and GISS, in response to the first JOC call for tropical wind experiments, revealed discrepancies which seemed to have been the result of differences in grid resolution but may also have reflected differences in more physical properties of the models. The question of model-dependence, and other factors affecting the comparison of simulation study results obtained by the various groups, received considerable attention at the October meeting. Closer cooperation is planned among simulation study groups to minimize the effect of the above factors in the future.

Toward this end, participants in the October 12th meeting established a standard horizontal grid resolution of 5° (latitude and longitude) and a minimum of 6

vertical levels, for subsequent numerical experiments. A hypothetical standard FGGE observing system was also established, with specified error limits on all firm elements of the observing system. This standard observing system will provide a common baseline for measuring the effect of adding or subtracting individual subsystems being considered for FGGE.

It is hoped that as a result of the establishment of uniform specifications in grid resolution and hypothetical FGGE systems, the effect of model dependence (although not the incompatibility or rejection effect) will be clearly exhibited when results from different modeling groups are compared. The resultant indication of the magnitude of one important variety of model-dependence in the simulation studies should be useful in the evaluation of these studies in future planning exercises.

References

- Charney, J., M. Halem, and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere. J. Atmos. Sci., 26, 1160-1163.
- Jastrow, R., and M. Halem, 1970: Simulation studies related to GARP. Bull. Amer. Meteor. Soc., 51, 490-513.
- 3. Williamson, D., and A. Kasahara, 1971: Adaptation of meteorological variables forced by updating. J. Atmos. Sci., 28, 1313-1324.
- Kasahara, A., 1972: Simulation experiments for meteorological observing systems for GARP. Bull. Amer. Meteor. Soc., 53, 252.
- Williamson, D. L., and R. E. Dickinson, 1972: Periodic updating of meteorological variables. J. Atmos. Sci., 29, 190-193.
- Jastrow, R., and M. Halem, 1971: Asynoptic assimilation of satellite temperature data. Princeton Meeting, Joint Organizing Committee, 1969.
- 7. Smagorinsky, J., 1972: Private communication.
- Morel, P., G. Lefevre, and G. Rabreau, 1971: On initialization and non-synoptic data assimilation. Tellus, 23, 197-206.
- 9. Morel, P., 1972: Private communication.
- Kasahara, A., and D. Williamson, 1972: Evaluation of tropical wind and reference pressure measurements: Numerical experiments for observing systems. *Tellus*, 24, 100.
- Cordon, C. T., L. Umscheid, Jr., and K. Miyakoda: Simulation experiments for determining wind data requirements in the tropics. To be published in J. Atmos. Sci.
- 12. Miyakoda, K., 1972: Private communication.
- Baumhefner, D. P., and P. R. Julian, 1972: The reference-level problem: Its location and use in numerical weather predictions. J. Atmos. Sci., 29, 285-299.
- 14. Smagorinsky, J., K. Miyakoda, and R. F. Strickler, 1970: Relative importance of variable in initial conditions for dynamical weather prediction. *Tellus*, 22, 141-157.